DOCUMENT RESUME

ED 466 600 IR 021 335

| | |
|---|---|
| AUTHOR | Karoulis, Athanasis; Pombortsis, Andreas |
| TITLE | Heuristic Evaluation of Web-Sites: The Evaluators' Expertise and the Heuristic List. |
| PUB DATE | 2001-00-00 |
| NOTE | 7p.; In WebNet 2001: World Conference on the WWW and Internet Proceedings (Orlando, FL, October 23-27, 2001); see IR 021 310. |
| PUB TYPE | Reports - Research (143) -- Speeches/Meeting Papers (150) |
| EDRS PRICE | EDRS Price MF01/PC01 Plus Postage. |
| DESCRIPTORS | *Evaluation Methods; Evaluators; *Heuristics; Knowledge Level; *Prior Learning; Surveys; *Training; *World Wide Web |
| IDENTIFIERS | *Expertise; *Web Sites |

ABSTRACT

Research on the evaluation of World Wide Web sites has already begun, but it is proceeding at a slow rate. The main reasons for this are the attempt to adapt existing methodologies to the particularities of the Web, the individual structure of Webs sites, and the issue of finding the appropriate evaluators. This study addresses these points and suggests a heuristic approach to the evaluation of Web sites. In the study, the evaluators were trained in the particularities of the heuristic evaluation in its classic form as well as in its Web-adapted form. Next, the researchers used Web-adapted heuristics, found in relative literature, and clarified them to the evaluators. Finally, the evaluators were involved in a real evaluation of five Web sites, and they wrote down their comments on questionnaires. The results confirmed two known conclusions: that the method is applicable to the Web; and that the prior evaluators' expertise is of great importance. It was also concluded that it is possible to augment this expertise in a short way, so that the evaluators have an increased performance during the evaluation as well. The main conclusion was that the heuristic list used performed inadequately, but the trend of the evaluators following a somewhat similar mode of thinking was noted, thus providing a way to adapt these heuristics in a more holistic approach to the Web. (Contains 27 references.) (Author/MES)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 466 600

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

G.H. Marks

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

# Heuristic Evaluation of Web-Sites: The Evaluators' Expertise and the Heuristic List

Karoulis Athanasis and Pombortsis Andreas
*Department of Informatics – Aristotle University of Thessaloniki*
*PO Box 888 - 54600 Thessaloniki - Greece*
{karoulis, apombo}@csd.auth.gr

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

**Abstract**: Research for the evaluation of websites has already begun, however it is proceeding at a very slow rate. The main reasons for this are, in our opinion, the attempt to adapt existing methodologies to the particularities of the web, the individual structure of web-sites and the issue of finding the appropriate evaluators. This study copes exactly with these points and suggests a heuristic approach for the evaluation of web sites.

In our study we tried primarily to train the evaluators in the particularities of the heuristic evaluation, in its classic form as well as in its web-adapted form. By doing this we try to answer the core question if we can augment the evaluators' expertise with a kind of training prior to the conduction of the evaluation itself. Next we used web-adapted heuristics, found in relative literature and tried to clarify them to the evaluators as well. Finally the evaluators were involved in a real evaluation of five web sites and they wrote down their comments on appropriately prepared questionnaires.

The results from this study confirm firstly two known conclusions, that the method is applicable to the Web and that the prior evaluators' expertise is of great importance. Yet, in addition to these, we concluded that it is possible to augment, under conditions, this expertise in a short way so they have an increased performance during the evaluation as well. Our main conclusion is, however, that the used heuristic list performed inadequately, but we noted the trend of the evaluators following a somewhat similar mode of thinking, thus providing us with the way to adapt these heuristics in a more holistic approach to the web.

## Keywords

Evaluation, heuristic evaluation, human computer interface, web, web evaluation.

## Introduction

Maybe the most frequently encountered evaluation method, of any entity, is the provision of a list of criteria (heuristics) relative to this entity followed by questioning in order to express peoples' opinions. These people can be users or experts on the particular domain. So we distinguish between user-based evaluations, known as "empirical evaluations" and expert-based evaluations. What can we evaluate in this way? Makrakis (1999) says everything has to do with:

- The design
- The organization
- The function
- The result of the entity under consideration

However, a number of problems arise from this approach.

- It provides all the disadvantages of the expert-based evaluations (Karat et al., 1992; Nielsen, 1993a; Karoulis et al., 2000b).
- The axes and heuristic list may become very long (Lewis & Rieman, 1994; Nielsen, 1993a). For example, the full interface usability criteria list suggested by Smith & Mosier (1986) includes 944 criteria.
- The evaluators' expertise plays a major role. (Lewis & Rieman, 1994; Nielsen, 1993b). We discuss this issue in detail later.

IR021335

BEST COPY AVAILABLE

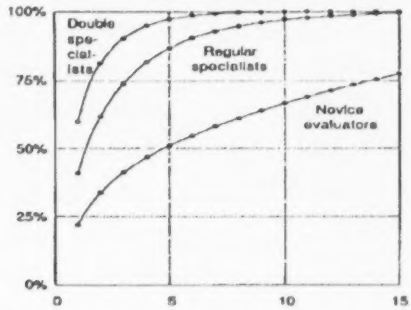
## The Heuristic Evaluation

To handle these problems Jacob Nielsen and Rolf Molich started their research in 1988 and in 1990 they presented the "heuristic evaluation" (Nielsen & Molich, 1990). The basic point was the reduction of the set heuristics to just a few, at the same time being broadly applicable and generally agreed; simultaneously augmenting the evaluators' expertise, and consequently their reliability. The method refers mainly to human-computer interface evaluation, yet a number of studies (Nielsen & Norman, 2000; Instone, 1997; Levi & Conrad, 1996) have proven its easy adaptability to the evaluation of web sites as well. This study belongs to this category.

The appropriate number of evaluators and their expertise are an issue of great importance. Researches up to now (Nielsen & Molich, 1990; Nielsen, 1992; Nielsen, 1993b) have shown that:

1. *Simple or novice evaluators.* They do not perform very well. We need 15 evaluators to find out 75% of the heuristically identifiable problems. The research has shown that 5 of these simple evaluators can pinpoint only 50% of the total problems.
2. *HCI experts (regular specialists).* They perform significantly better: 3 to 5 of such evaluators can point out 75% of the heuristicaly identifiable problems and among them all major problems of the interface.
3. *Double experts (specialists).* These are HCI experts with additional expertise on the subject matter. The reasearch has shown that 2-3 of them can point out the same percentage as the HCI experts.

The following diagram by Nielsen (1992) summarizes these statements.




As we can see in the diagram, to point out 75% of the heuristically identifiable problems we need 15 simple evaluators, while 3 expert evaluators bring the same result.

## Adaptation to the Web

Evaluation in the web differs from the traditional evaluation methodologies in many ways, due to the particularities of the web: every web site is an information space with non-linear structure, so two parameters, the download time and the ease of navigation, are of great importance. In addition to this, the evaluation procedure can be conducted by every evaluator on his/her own, redefining the notion of the "evaluation session" and introducing the notion of the "asynchronous evaluation", since the evaluators can perform their work from different places and at different time. Finally, in the web every evaluator is at the same time a user. Norman (2000) presents, for example, a cognitive walkthrough (Wharton et al., 1992; Lewis et al., 1990; Karoulis et al, 2000) performed in the web, playing the role of the simple user and thus proving the efficiency of this combination. This particular occurance on its own adds to the expert based evaluations in the web the hue of the empirical evaluation as well, augmenting its reliability, since the combination of user-based and expert-based approaches seems to provide the best results (Karat et al., 1992; Karoulis & Pombortsis, 2000; Karoulis et al., 2000b). The adaptation of the heuristic evaluation in the web has been already studied by researchers (eg. Instone, 1997; Levi & Conrad, 1996) and the results are in agreement that, in general, it is effective. Other researchers however consider that this issue has not yet been researched enough (Trochim, 1996; Lowe, 1999), and we adopt that opinion too.

## Research Questions

Given the youth of the web technology and the speed at which the web is growing, it seems indispensable that researchers conduct steady studies on its parameters, which are more often than not very variable. Bearing this in mind, in this study we are concerned with the specialized adaptation issues of the heuristic evaluation to the web and our main questions are as follows:

1. Can we, mainly, apply the heuristic methodology in the web?
2. Can the power users' expertise be augmented through some kind of "training", so that they can perform as well as expert evaluators?
3. Is the same list of heuristics valid for the web as for the evaluation of traditional interfaces?

The first question, as already mentioned, is a point of disagreement, consequently one more piece of evidence will strengthen either one of theses views.

Let us now consider the second question. It is known that it is possible for computer scientists to easily learn the evaluation methodologies and apply them sucessfully (Nielsen, 1992a; Wright and Monk, 1991). But computer scientists (the "experts") are not yet available in great numbers, so one can't argue that he/she will find someone to conduct the evaluation. So the following question arises; can some power users be trained in heuristic evaluation and be allowed to play the role of the expert? So our question refers in particular to how far a short training period can help the evaluators cross this zone in a feasible time towards the application of the heuristic evaluation methodology.

The heuristics are already broadly known and agreed, in the way Nielsen (1994b) suggests them. These heuristics have been adapted and ommented by Instone (2000) for their application in web-based heuristic evaluations, as follows:

1. *Visibility of system status*. The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.
2. *Match between system and the real world*. The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms.
3. *User control and freedom*. Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.
4. *Consistency and standards*. Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.
5. *Error prevention.* Even better than good error messages is a careful design which prevents a problem from occurring in the first place.
6. *Recognition rather than recall.* Make objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another.
7. *Flexibility and efficiency of use* Accelerators - unseen by the novice user - may often speed up the interaction.
8. *Aesthetic and minimalist design.* Dialogues should not contain information which is irrelevant or rarely needed.
9. *Help users recognize, diagnose, and recover from errors.* Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.
10. *Help and documentation.* Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation.

So the third question set in this study is if these heuristics are appropriate for the web and really efficient in the way Instone (2000) declares them to be.

## Adaptation, Organization and Conduction of the Evaluation

The evaluators conducted the evaluation in their own environment and using their own mode of internet connection. The first immediate consequence is the need to train the evaluators in a written manner in all necessary detail to fully clarify the procedure, but not to an excessive degree, so that phenomena of discouragement occure. Therefore we prepared a booklet, which we titled "Notes to the Evaluators", consisting of 7 pages, and describing the methodology of the heuristic evaluation by Nielsen (1992; 1994a and 1994b), as well as its adaptation in the web, and finally the description of the procedure the evaluators had to follow to complete their work. In addition to this, they have been equiped with another booklet, consisting of 5 pages,

containing the web-adapted heuristics of Instone (2000) and their comments. The training material included an "Evaluator's Notebook" as well, where the participants could note down their assessments and opinions.

## Results and Conclusions

Before starting the presentation of the results of this study, we would like to emphasize that the object of this study was not to evaluate the web sites under consideration, but to answer our research questions on the efficiency of the method and the chosen evaluators on the web. With this view in mind, we omit the results of the evaluations concerning the usability of each particular site. A direct consequence is that the aggregation of the evaluators' opinions is no longer necessary, as suggested by Nielsen (1994b) in order to obtain the evaluation results about the sites.

To proceed, we needed to group and categorize all the opinions of the evaluators in a separate supplementary document. In this document we provide the evaluators' opinions, as well as our assessments about the different heuristically identifiable problems. This approach can be found in the relevant literature (Nielsen, 1992, Lewis & Rieman, 1994) and relies on the observation that the conductors, in this case us, who are obviously HCI experts, can point out the majority of the usability problems during the preparation phase of the evaluation, that the evaluators will discover later on. This fact is a consequence of the discovery that a few double experts can pinpoint most of the problems. However, in our study this issue has proved to be insignificant, because the problems that were not discovered by us, but were found by the evaluators during the session, could be rated on their severity afterwards and matched with the corresponding heuristics. We noted our opinions in columns, next to those of the evaluators'.

The answers to our research questions can be given briefly as follows:

1. *Can we, mainly, apply the heuristic methodology in the web?*
   The answer to this question is affirmative, which is in agreement with most of the studies up to now. However, in order to apply the method effectively, the results of the following points must be taken into consideration as well.
2. *Can the power users' expertise be augmented through some kind of "training", so that they can perform as well as expert evaluators?*
   Yes and no. This also confirms the results from previous studies, that report the experts performing very differently from the simple users. However, this question is more complicated and will be discussed in detail later.
3. *Is the same list of heuristics valid for the web as for the evaluation of traditional interfaces?*
   The answer, according to our study, is negative. The heuristics we used seemed not to facilitate the evaluators in their work. They stated that they "interpreted" them to be applicable in different instances, and they provided us with some hints as well.

In more detail, heuristic evaluation performs well even in the web, yet the main issues of the evaluators' expertise and the validation of the web-heuristics remain.

The starting point for this study was the question if we could involve only power users, eg computer science students, instead of the difficultly in finding HCI experts. The separating line between these groups is not clear and our study can finally only approve the results of former studies (Nielsen & Molich, 1990; Nielsen, 1992) that suggest careful selection of the evaluators.

According to the mode of the training, four of the "successful" evaluators considered the booklet as "very lucid and enlightening", three considered a face-to-face seminar as a better solution without an optional booklet, while the rest had no opinion about this issue. There is additional evidence on this issue by Nielsen & Mack (1994), that heuristic evaluation can be taught in a half-day seminar, so this proposed approach seems to be a better one.

Regarding the third question on the appropriateness of the heuristics, it was clear that the used heuristics did not even facilitate the "successful" evaluators. On the contrary, some suggestions were made to us, as well as our collecting the evaluators' comments which resulted in a more lucid web-adapted heuristic list that seems to be more familiar and appears to facilitate the procedure.

Collecting the evaluators' answers we distinguished 28 categories of discovered problems. Most of these categories adhere to one or many of the above mentioned heuristics, as already commented. However, in this list there are categories that refer to the content of the web-site, its design as regards functionality or for not supporting the task the user wants to perform. These issues are obviously the concern of the user-centered design, which in the evaluators' opinion has not been applied, but is unavoidable, especially if one is designing for the web (Lewis & Rieman, 1994). Let us mention at this stage that in the assembling of this list we took into



BEST COPY AVAILABLE

consideration the results of this study and the evaluators' comments. However, the heuristics of Nielsen (2000b) and Instone (2000), as well as the prososals of Lowe (1999), which have been mentioned earlier, still remain as an underlying structure. Finally we also took into consideration the work of Togniazzini (2000), who proposes criteria (heuristics), not for the evaluation process itself, but for the design of web-sites, which are very close to the results of this study. The structure of the list we present is slightly different than usual: it consists of axes, which contain criteria (heuristics) as follows:

***Axis 1: Visible system status and in correspondence to what the user expects.***

*1.1.*: Navigation. Is it obvious where am I and where can I go next?
*1.2.*: Are all the icons and/or navigation possibilities visible and is it clear where they lead?
*1.3.*: Are all semantics clear and all functional graphics clear as to what they do?
*1.4.*: Is consistent language used, are international standards respected?

***Axis 2: Flexibility of use and structural integrity***

*2.1.*: Are there the necessary "accelerators" available? Can all pages be bookmarked?
*2.2.*: Has the site been debugged? Are there any empty areas or dangling and dead links?
*2.3.*: Does the site follow the conventions of the web?
*2.4.*: Does the site support its exploration? Is there a site map, search function etc.?
*2.5.*: Can the user easily remember the stucture, the functional and navigational mode of the site?

***Axis 3: Efficiency of use.***

*3.1.*: Are the technologies wisely used? Are these technologies acceptable for all user configurations?
*3.2.*: Are the response times of the site in line with what the user expects?
*3.3.*: Does the site adhere to the independent philosophy of the web?
*3.4.*: Does the site provide direct access to the most common tasks one can perform in it?

***Axis 4: User control, user-centered design and interaction***

*4.1.*: Can the user completely control all the interactive elements?
*4.2.*: Are there the corresponding interaction elements to the tasks that the user aims to perform?
*4.3.*: Is the feedback of these interaction elements of the kind the user expects?
*4.4.*: Does the site support all the tasks the user aims to perform?
*4.5.*: Can the user perform the tasks of his/her interest with minimal cognitive load?

***Axis 5: Content and presentation***

*5.1.*: Is there the right amount of information in the site (not insufficient or excessive)?
*5.2.*: Is there the right quality of information in the site (valid, clear, apropos)?
*5.3.*: Does the site give the impression of having been constructed and then left on its own?
*5.4.*: Is the information presented in a web-centric way, or is it just an adaptation of printed material?
*5.5.*: Is the information presented graphically acceptable? Easy to read?

***Axis 6: Subjective satisfaction, communication and help***

*6.1.*: Does the user feel he/she is isolated or left on his/her own?
*6.2.*: Is the site, in general, pleasant to use? Encourages exploration?
*6.3.*: Is there help, search function, external help, glossary?

The approach of building the list in axes containing criteria (heuristics) supports its application in two forms, as it may be obvious. One is the compact form – only the axes – if there is a shortage of resources (time, money etc.) or if we have very experienced evaluators available. The other one is the analytic form – all the criteria – for a more detailed evaluation of the site.

Summarizing the above, we argue that firstly special care must be taken in carefully selecting the evaluators, so that they have the nessesary expertise in computer science. Secondly, one has to follow a training approach with a seminar in addition to the booklet and finally use our proposed list of criteria that seems more familiar to the particular evaluator category. As a final conclusion to all the above, we believe that the method will finally have enough potential to provide an alternative solution in a situation where there are not HCI experts available to perform the evaluation.

## References


Instone, Keith. (1997) *Site Usability Evaluation*
http://www.webreview.com/1997/10_10/strategists/10_10_97_1.shtml 10 Oct.1997

Instone, Keith (2000). *Usability Heuristics For The Web.*
http://www.webreview.com/1997/10_10/strategists/10_10_97_2.shtml

Karat, C., Campbell, R. & Fiegel, T (1992). Comparison of Emperical Testing and Walkthrough Methods in User Interface Evaluation, 1992. *Proceedings of ACM CHI '92. Monterey, CA, May 3-7.*

Karoulis, A., Demetriades, S., Pombortsis, A (2000). The Cognitive Graphical Jogthrough – An Evaluation Method with Assessment Capabilities. *Applied Informatics 2000 Conference Proceedings,* February 2000, Innsbruck, Austria.

Karoulis, A., Demetriades, S., Pombortsis, A (2000b). Evaluation of multimedia educational interfaces for the junior highschool classes using two different methodologies: the "Perivallon" experience.

Karoulis, A., and Pombortsis, A (2000). Evaluating the Usability of Multimedia Educational Software for Use in the Classroom Using a "Combinatory Evaluation" Approach. *Proc. of Eden 4th Open Classroom Conference,* Barcelona, Spain, November 2000.

Levi, M.D., and Conrad, F.G (1996). A Heuristic Evaluation of a World Wide Web Prototype, *Interactions Magazine*, July/August, 1996. Vol.III.4, pp. 50-61. http://stats.bls.gov/orersrch/st/st960160.htm

Lewis, C., Polson, P., Wharton, C. & Rieman, J (1990). Testing a Walkthrough methodology for Theory-Based Design of Walk-Up-and-Use Interfaces. *Proceedings of ACM CHI '90, Seattle, Washington.* April 1-5, 1990, pp. 235-242.

Lewis, C. and Rieman, J (1994). *Task-centered User Interface Design - A practical introduction,* 1994. ftp.cs.colorado.edu/pub/cs/distribs/HCI-Design-Book.

Lowe, David (1999). Web Site Evaluation. *WebNet Journal,* vol.1, No.4, Oct-Dec, 1999.

Makrakis, B (1999). Evaluation of Open and Distance Learning Systems. In *Open and Distance Learning*, Vol A', Greek Open University, Patras, Greece, 1999. pp.245-301

Nielsen, Jacob (1992). Finding Usability Problems through Heuristic Evaluation. *Proceedings of ACM CHI '92* Monterey, CA, May 3-7, 1992

Nielsen, J (1992a). Evaluating the Thinking Aloud Technique for Use by Computer Scientists. In Hartson, H.R. and Hix, D. (Edts.) *Advances in Human-Computer Interaction,* vol. 3 Ablex, Norwood, New Jersey, 1992. pp. 69-82.

Nielsen, Jacob (1993a). Usability evaluation and inspection methods. In *Bridges between worlds*, INTERCHI '93 Tutorial notes 22. (Ashlund, S., Mullet, K., Henderson, A., Hollnagel, E. and White, T. eds.) Reading, MA: Addison-Wesley, 1993.

Nielsen, Jacob (1993b). *Usability Engineering*. Academic Press, San Diego, 1993.

Nielsen, Jacob(1994a). Enhancing the Explanatory Power of Usability Heuristics. *Proc. ACM CHI'94 Conf.* (Boston, MA, April 24-28), 152-158.

Nielsen, Jacob(1994b). Heuristic evaluation. In Nielsen, J., and Mack, R.L. (Eds.), *Usability Inspection Methods*, John Wiley & Sons, New York, NY.

Nielsen, J. and Mack, R.L. (edts.) (1994) *Usability Inspection Methods*. John Wiley & Sons, 1994.

Nielsen, J., and Molich, R (1990). Heuristic Evaluation of User Interfaces, *Proc. of Computer-Human Interaction Conference* (CHI), Seattle, WA, 1-5 April, 1990, p. 249-256.

Nielsen, J. and Norman, D (2000). *Web-site Usability: Get the Right Answers From Testing* http://www.useit.com 14/2/2000.

Norman, D.A (1988). *The Psychology of Everyday Things*. New York: Basic Books, 1988.

Norman, D (2000). Web-site Usability: Walk-Through: A Usability Experiment. http://www.informationweek.com/shared/infoweek/773 14/2/2000

Smith, S and Mosier, J (1986). Design Guidelines for Designing User Interface Software. The MITRE Corp., 1986. ftp://ftp.cis.ohio-state.edu/pub/hci/Guidelines

Togniazzini, B (2000). *Ask Tog: First Principles*. http://www.asktog.com/basics/firstPrinciples.html ,

Trochim, W (1996). *Evaluating Websites*. Cornell University, Ithaca, NY, http://trochim.human.cornell.edu/webeval/intro.htm (30/11/00), November, 1996.

Wharton, C., Bradford, J., Jeffries, R. & Franzke, M (1992). Applying Cognitive Walkthroughs to More Complex User Interfaces: Experiences, Issues and Recommendations, 1992. *Proceedings of ACM CHI '92. Monterey, CA, May 3-7.*

Wright, P.C. and Monk, A.F (1991). A Cost-Effective Evaluation Method for Use by Designers. *International Journal of Man-Machine Studies*, 35, 6 (Dec), 1991, pp.891-912.



BEST COPY AVAILABLE

**U.S. Department of Education**
*Office of Educational Research and Improvement (OERI)*
*National Library of Education (NLE)*
*Educational Resources Information Center (ERIC)*

# NOTICE

## Reproduction Basis

☒ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").